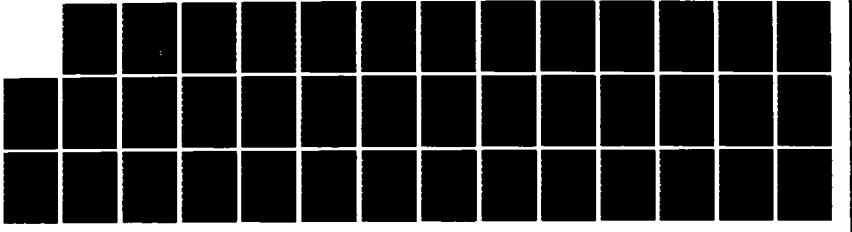
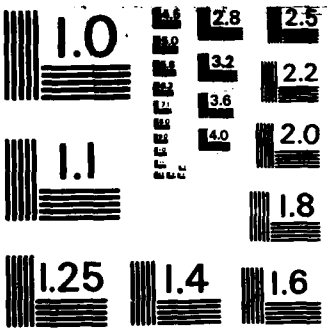
AD-A139 841 ESTIMATING NETWORK RELIABILITY WITH ACCELERATED CONVERGENCE RATES ON ERRO..(U) NORTH CAROLINA UNIV AT CHAPEL HILL CURRICULUM IN OPERATIONS R.. G S FISHMAN 1/1

UNCLASSIFIED DEC 83 UNC-ORSA-TR-83/7 N00014-76-C-0302 F/G 12/1 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS - 1963 - A

OPERATIONS RESEARCH AND SYSTEMS ANALYSIS

AD A139841

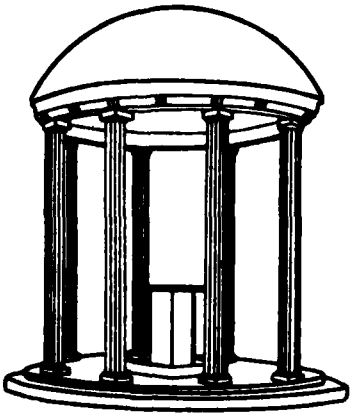
ESTIMATING NETWORK RELIABILITY WITH
ACCELERATED CONVERGENCE RATES ON ERROR BOUNDS

George S. Fishman

Technical Report No. UNC-ORSA-TR-83/7

December 1983

UNIVERSITY OF NORTH CAROLINA
AT CHAPEL HILL



ESTIMATING NETWORK RELIABILITY WITH ACCELERATED CONVERGENCE RATES ON ERROR BOUNDS

George S. Fishman

Technical Report No. UNC-ORSA-TR-83/7

December 1983

Curriculum in Operations Research
and Systems Analysis

University of North Carolina at Chapel Hill

This research was supported by the Office of Naval Research under contract N00014-76-C-0302. Reproduction in whole or in part is permitted for any purpose of the United States government.

Acknowledgment

I would like to express my gratitude to my colleague, Professor Scott Provan, for many illuminating conversations on reliability and to Mr. Tien-Yi Shaw for his considerable assistance in programming.

Abstract

Let $G=(V,E)$ denote an undirected network with node set V and arc set $E=\{1,\ldots,N\}$. Arcs fail randomly and independently with probability $1-q_i$ for $\forall\ i\in E$. This paper describes a numerical procedure for estimating $g(s,t)$, the probability that nodes s and t are connected for specified $s,t\in E$, with bounds on absolute error proportional to $1/K$ for a specified nonrandom finite sequence and proportional to $\log K/K$ for certain nonrandom infinite sequences, where K is the number of replications. These convergence rates are best possible. Although the infinite sequences have a slower convergence rate than the finite sequence has, they offer the convenience of allowing one to add replications and retain the rate whereas the finite sequence does not. These bounds improve on the convergence rate $O(1/K^{1/2})$ for the standard error in the case of independent Monte Carlo replications based on random sampling. Moreover, they hold with certainty. Algorithms for computing estimates are shown to have an upper bound $O(N)+O(\max(N,|V|))$ on time complexity per replication as $K\to\infty$.

The paper first describes the estimation of $g(s,t)$ for $q_1=\ldots=q_N=p$ by using a tabled truncated binomial distribution together with the A-canonical representation of a positive integer. It then describes how to incorporate stratified sampling to estimate $g(s,t)$ as a function of p at small marginal increase in time complexity. Next, the paper extends the stratified sampling method to the case of unequal $q_1,\ldots,q_N$.

An example based on a network of 30 arcs illustrates the techniques. Lastly, the paper extends the method to the estimation of $g(s,T)$, the probability that s and t are connected for all $t\in T\subseteq E-\{s\}$.

KEY WORDS: Monte Carlo methods, network connectedness, network reliability, quasirandom points.

Introduction

Let $G=(V,E)$ denote an undirected network with node set V and arc set $E=\{1,\ldots,N\}$. Arcs are assumed to fail randomly and independently with probability $1-q_i$ for all $i\in E$. We wish to estimate

$g(s,t)$ = the probability that nodes s and t are connected for specified $s,t\in V$.

It is well known that direct evaluation of $g(s,t)$ takes $O(2^{|E|})$ steps, a computation that can be infeasible even with a moderate number of arcs. At least three approaches have been suggested to reduce the severity of this difficulty. The first concentrates on networks with special structure. For example, see Rosenthal (1977). The second relies on finding bounding inequalities. For example, see Ball and Provan (1983) and Zemel (1982).

The third uses estimation techniques among which the most common are Monte Carlo procedures based on independent trials or replications that use random sampling. Van Slyke and Frank (1972) describe how Monte Carlo methods using a stratified sampling design lead to estimators of network reliability with considerably smaller variances than would obtain if pure random sampling were employed. Easton and Wong (1980) demonstrate a similar benefit using conditional sampling and Kumamoto, Tanaka and Inoue (1977) show a like benefit for dagger sampling, a special form of antithetic variates (Hammersley and Handscomb 1964). Karp and Luby (1983) describe a Monte Carlo procedure based on importance sampling whose computation time complexity to achieve a specified accuracy is linear in the number of cutsets.

While each exploits network structure in a specialized way, these Monte Carlo procedures share one shortcoming in common. They all lead to standard errors that converge as $O(1/K^{1/2})$ where K is the number of independent trials or replications. To improve this convergence rate Fishman (1983) describes a numerical procedure that uses quasirandom point sequences together with a conditionality property based on cutsets to produce an estimate of $g(s,t)$ whose absolute error has an upper bound proportional to $(\log K)^{|P-R|}/K$ where P denotes the union of all minimal s-t pathsets and R is a minimal cutset with the property that $|R \cap S| = 1$ for each s-t pathset S. Moreover, since these sequences are nonrandom, the bound holds with certainty. These results imply that for a given network G quasirandom point sequences produce a deterministic upper bound on absolute error that converges faster than $1/K^{1/2}$.

The purpose of this paper is to describe related estimation techniques that improve ~~this~~ the convergence rate to $1/K$ when a specified nonrandom finite sequence of sample points is used and to $(\log K)/K$ when certain nonrandom infinite sequences of sample points are used. In addition, algorithms are given for computing estimates of $g(s,t)$ with $1/K$ and $(\log K)/K$ convergence rates with computation time complexities per replication having an upper bound $O(N) + O(\max(N,|V|))$ as $K \to \infty$. Although the infinite sequences have a slower convergence rate than the finite sequence has, ~~they~~ these offer the convenience of allowing one to add replications as desired and retaining the rate. By contrast, the finite sequence does not allow this addition, once K is initially fixed.

Section 1 describes the technique for the special case of equal failure probabilities. In particular, it shows how sampling from a tabled truncated binomial distribution and applying the A-canonical representation of an integer B (Kruskal 1963 and Katona 1966) make these convergence rates and time complexity possible. Section 2 then shows how to incorporate stratified sampling to allow the estimation of $g(s,t)$ as a function of p at relatively small increase in time complexity while retaining the desirable convergence rates. Section 3 shows how the method of stratified sampling facilitates the estimation of $g(s,t)$ for the case of unequal $q_1,\ldots,q_N$, while preserving the convergence rates and time complexity. Section 4 illustrates the technique of Section 1 for a network of 30 arcs. Lastly, Section 5 extends the analysis to the case of estimating $g(s,T)$, the probability that s is connected to all arcs in $T \subseteq E-\{s\}$.

1. The Basic Method

For all nodes $u,v \in V$ define:

$C(u,v)$ = minimal u-v cutset of minimal cardinality

$M(u,v)$ = minimal u-v pathset of minimal cardinality

$L(u,v)$ = set of all minimal u-v pathsets

$P(u,v)$ = union of all minimal u-v pathsets

and

$V(u,v)$ = node set corresponding to $P(u,v)$.

Let

$$F_j(n,\theta) = \sum_{i=0}^{j} \binom{n}{j} \theta^j (1-\theta)^{n-j} \qquad (1)$$

$$n = \text{integer} > 0 \quad \text{and} \quad 0 < \theta < 1$$

$$0 \le j \le n .$$

For each arc $i=1,\ldots,N$ let

q_i = probability that arc i operates

and

$z_i = 1$ if arc i operates

$= 0$ if arc i fails.

Then for the network, let

$z = (z_1,\ldots,z_N)$ = network state

Z = set of all network states z

Z_m = set of all network states with $z_1+\ldots+z_N=m$

$$k(u,v|z) = 1 - \prod_{S \in L(u,v)} \left(1 - \prod_{i \in S} z_i\right) \qquad (2)$$

$$R_m(u,v) = \sum_{z \in Z_m} k(u,v|z) \qquad (3)$$

and

$$R(u,v) = \sum_{z \in Z} k(u,v|z) = \sum_{m=1}^{N} R_m(u,v) . \qquad (4)$$

Note that given z $k(u,v|z)$ is the conditional probability (0 or 1) that u and v are connected. Also $R_m(u,v)$ denotes the number of distinct states z that induce u-v connectedness when m arcs operate and $R(u,v)$ denotes the number of distinct states z that induce u-v connectedness regardless of the number of operating arcs.

Expression (2) is of seminal importance throughout this paper. Suppose we collect data $x_1,\ldots,x_K$ on K sampling experiments or replications where

$$x_j = (x_{1j},\ldots,X_{Nj}) \qquad j=1,\ldots,K$$

and

$$x_{ij} = 1 \quad \text{if arc } i \text{ operates on replication } j$$
$$= 0 \quad \text{otherwise.}$$

Here $x_1,\ldots,x_K$ are sample network states and

$$\hat{g}_K(s,t) = \frac{1}{K} \sum_{j=1}^{K} k(s,t|x_j) \tag{5}$$

is the proportion of trials on which s and t are connected. As we show shortly the sampling plan for choosing $x_1,\ldots,x_K$ determines how good an approximation $\hat{g}_K(s,t)$ is to $g(s,t)$.

It is important to note that (2) is used merely for representation and analysis. To determine the value of $k(u,v|x_j)$ on each replication, we assume that a "depth-first search" algorithm, as in Aho, Hopcroft and Ullman (1974, pp. 176-177), is employed. This algorithm has an upper bound on computation time complexity $O(|P(u,v)|, |V(u,v)|)$ where

$|S|$ denotes the cardinality of the set S .

Sampling Sequences and Error Bounds

We now turn to the selection of a sampling scheme for generating the data $x_1,\ldots,x_K$. In particular, we would like a scheme that enables us to estimate $g(s,t)$ by $\hat{g}_K(s,t)$ and related estimators in a computationally efficient manner. To make an informed selection, we first need to consider several basic ideas from the theory of equidistributed points. Let

$$I_{[\alpha,\beta)}(x) = 1 \quad \text{if} \quad \alpha \le x < \beta$$

$$= 0 \quad \text{otherwise}$$

and let $v_1, v_2,\ldots$ denote a sequence of points in $[0,1)$. Then the sequence $v_1, v_2,\ldots$ is said to be uniformly distributed or equidistributed on $[0,1)$ if (Schmidt 1977, p. 1)

$$\lim_{K\to\infty} \frac{1}{K} \sum_{j=1}^{K} I_{[\alpha,\beta)}(v_j) = \beta-\alpha \tag{6}$$

for all α and β such that $0\le\alpha<\beta\le 1$. Conversely, if $v_1, v_2,\ldots$ are equidistributed in $[0,1)$ then (6) holds. As a measure of error in estimating $\beta-\alpha$ one has the discrepancy measures

$$D_K = \sup_{0\le\alpha<\beta\le 1} \left|\frac{1}{K} \sum_{j=1}^{K} I_{[\alpha,\beta)}(v_j) - (\beta-\alpha)\right| \tag{7a}$$

and

$$D_K^* = \sup_{0 \le \alpha \le 1} \left| \frac{1}{K} \sum_{j=1}^{K} I_{[0,\alpha)}(v_j) - \alpha \right|, \tag{7b}$$

with the inequality (Kuipers and Niederreiter 1974, p. 91)

$$D_K^* \le D_K \le 2D_K^* \quad . \tag{8}$$

Note that D_K and D_K^* are <u>worst case</u> measures of the errors of estimation.

For the <u>finite sequence</u> $\{v_j = \frac{2j-1}{2K} \; ; \; j=1,\ldots,K\}$ it is known (Niederreiter 1978, p. 972) that

$$D_K^* = 1/K, \tag{9}$$

which is best possible. In this paper we devise a sampling scheme that enables us to write $\hat{g}_K(s,t)$ and related estimates as linear combinations of indicator functions with argument v_j and thus are able to realize the convergence rate $1/K$.

Although one can achieve this rate in practice for $\hat{g}_K(s,t)$ in (5), situations can arise in which after K replications, one finds that a larger sample size is necessary. Unfortunately the finite sequence $\{v_j = \frac{2j-1}{2K} \; ; \; j=1,\ldots,K\}$ does not allow for adding points $v_{K+1}, v_{K+2},\ldots,v_{K+J}$ and achieving a bound proportional to $1/(K+J)$ on discrepancy. However, it is known that there exist <u>infinite sequences</u> $v_1, v_2, \ldots$ (Kuipers and Niederreiter 1974, p. 125) such that

$$D_K \le c \, (\log K)/K \tag{10}$$

where c depends only on the sequence. One such sequence due to van der Corput (1935) considers the unique dyadic expansion of the integer n

$$n = \sum_{i=0}^{I} a_i 2^i \qquad a_i \in \{0,1\} \quad 0 \le i \le I$$

where $\lfloor I = \log_2 n \rfloor$. Then the radical inverse function

$$\phi_2(n) = \sum_{i=1}^{I} a_i 2^{-i-1}$$

can be used to compute the van der Corput sequence $\{v_j = \phi_2(n+j) ; j=1,2,\ldots\}$ for which unpublished results of Tijdeman (Niederreiter 1974, p. 973) show that for all $K \ge 1$

$$D_K^* \le (\frac{1}{3} \log_2 K + 1)/K \tag{11a}$$

and

$$\limsup_{K \to \infty} (KD_K^* - \frac{1}{3} \log_2 K) \ge \frac{4}{9} + \frac{1}{3} \log_2 3 . \tag{11b}$$

Bejian and Faure (1977) show that (11a) also holds for D_K and (11b) holds with equality for D_K .

Schmidt (1977, p. 28) shows that for infinite sequences the best possible lower bound is

$$D_K \ge O((\log K)/K)$$

so that the van der Corput sequence achieves the fastest convergence rate with respect to K . Moreover, no other infinite sequence is known that has uniformly smaller discrepancy than this sequence has

(Kuipers and Niederreiter 1974, p. 127). Note that the van der Corput sequence is one form of a quasirandom point sequence.

It is important to note that both the finite and infinite sequences discussed here are nonrandom so that the equality (9) and the bound (10) hold with certainty. By contrast, independent replications using random sampling lead to (Chung 1949)

$$\limsup_{K \to \infty} \frac{\sqrt{2K}\, D_K^*}{\sqrt{\log \log K}} = 1 \qquad \text{w.p. 1} ,$$

indicating a considerably slower rate of convergence. Therefore, provided that we can effectively control the time complexity of our proposed sampling plans, the foregoing finite and van der Corput sequences should be preferred to independent random sampling.

Canonical Representation

As we show shortly the sampling method that we adopt requires that we be able to use a single number v_j on replication j to select the network state z from Z or, as in Sections 2 and 3, from Z_m. To make this selection possible in a relatively efficient manner we rely on an extremely useful mathematical relationship. Let B be a positive integer. Then there exists an A-canonical representation of B (Kruskal 1963 and Katona 1966)†

$$B = \binom{B_A}{A} + \binom{B_{A-1}}{A-1} + \ldots + \binom{B_\ell}{\ell} \tag{12}$$

where

$$B_A > B_{A-1} > \ldots > B_\ell \geq \ell \geq 1 .$$

†I am grateful to my colleague, Professor Scott Provan, for making me aware of this relationship.

Most importantly, this representation is unique.

To determine $B_A, B_{A-1},\ldots,B_\ell$ for $B>0$ one can use the relationship (Ball and Provan 1983, p. 256)

$$B_A = \max[b: \binom{b}{A} \le B] \tag{13a}$$

$$B_i = \max[b: \binom{b}{i} \le B - \sum_{j=i+1}^{A} \binom{B_j}{j}] \qquad i=A-1,A-2,\ldots,\ell$$

where

$$\ell = \max[i: B = \sum_{j=i}^{A} \binom{B_j}{j};\ i\ge 1] . \tag{13b}$$

Suppose that we know that exactly A of N arcs operate on replication j so that $\sum_{i=1}^{N} x_{ij} = A$. To determine exactly which of the N arcs operate one can proceed as follows: Sample B from the discrete uniform distribution on $\{0,1,\ldots,\binom{N}{A}-1\}$. The quantity B is the "number" of the combination of A operating and $N-A$ failed arcs on replication j. Then if $B=0$ one takes

$$\begin{aligned} x_{ij} &= 1 \qquad i=1,\ldots,A \\ &= 0 \qquad \text{otherwise.} \end{aligned} \tag{14a}$$

If $B>0$ and $\ell=1$ one takes

$$\begin{aligned} x_{ij} &= 1 \qquad i=B_m+1 \qquad m=1,\ldots,A \\ &= 0 \qquad \text{otherwise .} \end{aligned} \tag{14b}$$

Lastly, if $B>0$ and $\ell>1$ one takes

$$\begin{aligned} x_{ij} &= 1 \qquad i=1,\ldots,\ell-1\ ,\ B_\ell+1,\ldots,B_A+1 \\ &= 0 \qquad \text{otherwise.} \end{aligned} \tag{14c}$$

Note that these assignments are unique functions of A and B.

Equal Probabilities

We first consider the case of s-t connectedness with equal operating probabilities $q_i=p$ $i\epsilon P(s,t)$. For convenience of exposition we take $C=C(s,t)$, $L=L(s,t)$, $M=M(s,t)$, $P=P(s,t)$, and $R_m=R_m(s,t)$, and assume $P=E$ so that $|P|=N$. Here the number of operating arcs has the binomial distribution with probability distribution function (1) with $n=N$ and $\theta=p$.

Algorithm C.1 describes how to conduct the K replications that lead to the computation of $\hat{g}_K(s,t)$ as an estimate of $g(s,t)$. Note that the number of operating arcs A is determined (step 2c) from a truncated distribution. To appreciate why this is so observe that if the number of operating arcs on a replication exceeds $N-|C|$, then s and t are connected with probability 1. Conversely, if fewer than $|M|$ arcs operate then with probability 1 s and t are not connected. Therefore, it is to our advantage to restrict sampling (step 2c) to the uncertain connectedness outcomes when the number of operating arcs exceeds $|M|-1$ but is less than $N-|C|+1$. Also note that A can be determined on each replication with time complexity $O(1)$ using a procedure described in Fishman and Moore (1981).

Theorem 1 gives the implications of algorithm C.1.

Theorem 1. Suppose algorithm C.1 is used to compute $\hat{g}_K(s,t)$. Then:

(i) For the finite sequence $\{v_j = \frac{2j-1}{2K}\ ;\ j=1,\ldots,K\}$

$$|\hat{g}_K(s,t) - g(s,t)| \le 2HQ/K \qquad (15a)$$

where

$$Q \le \min\{ \sum_{m=\lceil M \rceil}^{N-\lceil C \rceil} R_m \ , \ \sum_{m=\lceil M \rceil}^{N-\lceil C \rceil} [\binom{N}{m} - R_m] \} \ . \tag{15b}$$

(ii) There exist infinite sequences $\{v_j \ ; \ j=1,2,\ldots\}$ such that the bound in (15a) holds with $c(\log K)/K$ replacing $1/K$ and where c depends only on the selected sequence.

(iii) One can compute each iteration of step 2 with an upper bound $O(Np) + O(N) + O(\max(N,|V|))$ on time complexity per replication as $K \to \infty$.

Proof. For (i) observe that the number of operating arcs A has a truncated binomial distribution (step 2c) on the integers $|M|,\ldots,N-|C|$ with parameters N and p. Also, for given A it is clear that any of the $\binom{N}{A}$ combinations of A operating arcs are equally likely (step 2d). The quantity B merely denotes the number of the combination (given A) selected.

Let $m=A$ and suppose that the $\binom{N}{m}$ possible combinations of operating arcs are numbered $1,\ldots,\binom{N}{m}$. Let $J_M = \{1,\ldots,\binom{N}{M}\}$ denote the set of all these indices, let J_m^* denote the subset of J_m for which s and t are connected and, finally, let $F_m = F_m(N,p)$. Then the unique A-canonical representation of B leads to

$$\begin{aligned} k(s,t|x_j) &= \sum_{m=\lceil M \rceil}^{N-\lceil C \rceil} \sum_{\ell \in J_m^*} I_{[\frac{\ell-1}{\binom{N}{m}} \ , \ \frac{\ell}{\binom{N}{m}})} \left(\frac{Hv_j - F_{m-1} + F_{|M|-1}}{F_m - F_{m-1}} \right) \\ &= \sum_{m=\lceil M \rceil}^{N-\lceil C \rceil} \sum_{\ell \in J_m^*} I_{[a_{\ell m} \ , \ b_{\ell m})} (v_j) \end{aligned} \tag{16}$$

Algorithm C.1

Purpose: To compute $\hat{g}_K(s,t)$ as an estimate of $g(s,t)$.

Given: N, p, $|C|$, $|M|$, K , s and t.

1. Set parameters: n=N, θ=p and S=0.
2. Perform K replications:
   a. For i=1,...,n set x_{ij}=0.
   b. Select v_j .
   c. Determine the number of operating arcs:

$$A=\min\{m: [F_m(n,\theta)-F_{|M|-1}(n,\theta)]/H>v_j;\ |M|\le m\le n-|C|\}$$

   where $H \equiv F_{n-|C|}(n,\theta)-F_{|M|-1}(n,\theta)$.

   d. Determine the "number" of the combination of A operating and N-A failed arcs:

$$B = \left\lfloor \frac{Hv_j-F_{A-1}(n,\theta)+F_{|M|-1}(n,\theta)}{\theta^A(1-\theta)^{n-A}} \right\rfloor .$$

   e. Determine the A-canonical representation ℓ , $B_\ell,\ldots,B_A$ of B using (13). If B=0 , ℓ=A+1 .
   f. If ℓ>1 , then for v=1,...,ℓ-1 ; B_v=v-1 .
   g. Set operating indicators:

      For v=1,...,A ; set i=B_v+1 and x_{ij}=1 .

   h. Use a labeling algorithm to determine $k(s,t|x_j)$.

      ($k(s,t|x_j)$=1 if s and t are connected

      =0 otherwise.

   i. Accumulate results:

      S=S + $k(s,t|x_j)$.

3. Compute final estimate for K replications:

$$\hat{g}_K(s,t) = 1-F_{n-|C|}(n,\theta) + H\cdot S/K .$$

4. Done.

where H is defined in step 2c and

$$a_{\ell m} = [F_{m-1}-F_{|M|-1} + \frac{\ell-1}{\binom{N}{m}} (F_m-F_{m-1})]/H$$

$$= [F_{m-1}-F_{|M|-1} + (\ell-1)\, p^m(1-p)^{N-m}]/H \qquad (17a)$$

and

$$b_{\ell m} = a_{\ell m} + [p^m(1-p)^{N-m}]/H \; . \qquad (17b)$$

Now observe that one can write

$$g(s,t) = 1-F_{N-|C|}(N,p) + \sum_{m=|M|}^{N-|C|} \sum_{\ell \epsilon J_m^*} p^m(1-p)^{N-m} \; .$$

Therefore,

$$\hat{g}_K(s,t) - g(s,t) = 1-F_{N-|C|}(N,p) + \frac{H}{K} \sum_{j=1}^{K} k(s,t|x_j)$$

$$- 1 + F_{N-|C|}(N,p) - \sum_{m=|M|}^{N-|C|} \sum_{\ell \epsilon J_m^*} p^m(1-p)^{N-m}$$

$$\leq H \left| \sum_{m=|M|}^{N-|C|} \sum_{\ell \epsilon J_m^*} \left[\frac{1}{K} \sum_{j=1}^{K} I_{[a_{\ell m}, b_{\ell m})}(v_j) - p^m(1-p)^{N-m}/H\right]\right| \; .$$

Using the results in (7) and (9) it is clear that this bound is bounded above by $2HQ/K$ where Q is a function of the number of nonconnected intervals in $\{[a_{\ell m}, b_{\ell m}): \ell \epsilon J_m^* \; ; \; m=|M|,\ldots,N-|C|\}$. In particular, the worst possible arrangement of intervals leads to

$$Q \leq \min\{\sum_{m=|M|}^{N-|C|} R_m, \sum_{m=|M|}^{N-|C|} [\binom{N}{m} - R_m]\} \; ,$$

which establishes (i). The result (ii) follows immediately by using (10) instead of (7) in (15a).

For the result in (iii) note that steps 2a and 2b have time complexities $O(n)$ and $O(1)$, respectively, per replication. Also, step 2c can be performed with time complexity $O(1)$ per replication using the cutpoint sampling method described in Fishman and Moore (1981). There tables need to be computed (with time complexity $O(n)$ and space complexity $O(2n)$) prior to using algorithm C.1. Step 2d requires $O(1)$ time per replication. Also, the identification of operating arcs (steps 2e and 2f) takes $O(n)$ time. As $K\to\infty$, step 2g takes $O(n\theta)$ time on average, and step 2i takes $O(1)$ time. Since step 2h based on a depth-first algorithm takes $O(\max(N,|V|))$ time the result in (iii) obtains. This completes the proof of Theorem 1.

Observe that the bounds on convergence in (i) and (ii) are proportional to $H=F_{N-|C|}(N,p) - F_{|M|-1}(N,p)$. In particular, for fixed N, C and M this quantity decreases to zero as $p\to 0$ and $p\to 1$. Also if C and M are unknown one can substitute 0 for $|C|$, $|M|$ or both in algorithm C.1, which continues to compute correctly but with a different coefficient in the bounds in (i) and (ii).

The quantity Q is of principal interest here. It depends on the arrangement of subintervals in the unit interval assigned to the sample outcomes for which s and t are connected for each $m=|M|,\ldots,N-|C|$. If for given m these subintervals form one long interval with no breaks, then $Q=N-|C|-|M|+1$. However, the number of steps necessary to make this arrangement prior to sampling is comparable to the direct computation of $g(s,t)$ by total enumeration. Nevertheless, arrangements that are clearly preferable to others hopefully can be made at relatively small cost. These assignments are a topic of continuing research.

2. Stratified Sampling

We naturally would like to extend the techniques of Section 1 to the estimation of $g(s,t)$ when at least some of the probabilities $q_1,\ldots,q_N$ are distinct. This extension is possible if one exploits certain features of a sampling technique known as stratified sampling. Van Slyke and Frank (1972) propose this technique for estimating network reliability characteristics in an environment in which $q_1,\ldots,q_N$ are identical and samples are drawn randomly within strata. The effect there is to reduce the variances of the resulting estimators below the corresponding variances that would obtain for pure random sampling. This section retains the restriction $q_1=\ldots=q_N=p$ but replaces random sampling within strata with the nonrandom sequences used in Section 1 and demonstrates how stratified sampling affects the bounds on error convergence. Then Section 3 shows how the stratified sampling plan can be extended to the case of distinct $q_1,\ldots,q_N$.

Let

$$w_m(s,t) = \text{probability that } s \text{ and } t \text{ are connected given that exactly } m \text{ arcs are operating.} \tag{18}$$

Now observe that $w_m(s,t)$ can be written in the form

$$w_m(s,t) = R_m(s,t)/\binom{N}{m} = \frac{1}{\binom{N}{m}} \sum_{z\in Z_m} k(s,t|z) \tag{19}$$

so that

$$\begin{aligned} g(s,t) &= \sum_{m=0}^{N} w_m(s,t)\ \binom{N}{m} p^m(1-p)^{N-m} \\ &= 1-F_{N-|C|}(N,p) + \sum_{m=|M|}^{N-|C|} p^m(1-p)^{N-m} \sum_{z\in Z_m} k(s,t|z)\ . \end{aligned} \tag{20}$$

Algorithm C.2 describes a procedure for computing

$$\tilde{w}_m(s,t) = \frac{1}{K} \sum_{j=1}^{K} k(s,t|x_j) \qquad (21)$$

as an estimate of $w_m(s,t)$ where B in step 2c gives the "number" of the combination of $A=m$ operating arcs on replication j.

Algorithm C.2

Purpose: To compute $\tilde{w}_m(s,t)$ as an estimate of $w_m(s,t)$.

Given: N, m, K_m, s and t $(|M| \le m \le N - |C|)$.

1. Set parameters: $n=N$, $A=m$ and $S=0$.
2. Perform K_m replications:
   a. For $i=1,\ldots,n$ set $x_{ij}=0$.
   b. Select v_j.
   c. Determine the "number" of the combination of A operating arcs and $N-A$ failed arcs:

      $$B = \lfloor \binom{n}{A} v_j \rfloor .$$

   d. Determine the A-canonical representation ℓ, $B_\ell,\ldots,B_A$ of B using (13). If $B=0$, $\ell=A+1$.
   e. If $\ell>1$, then for $v=1,\ldots,\ell-1$; $B_v=v-1$.
   f. Set operating indicators:

      For $v=1,\ldots,A$; set $i=B_v+1$ and $x_{ij}=1$.

   g. Use a labeling algorithm to determine $k(s,t|x_j)$.
   h. Accumulate results:

      $S=S+k(s,t|x_j)$.

3. Compute final estimate for K_m replications:

   $$\tilde{w}_m(s,t) = S/K_m .$$

4. Done.

Let

$$\tilde{g}_K(s,t) = 1-F_{N-|C|}(N,p) + \sum_{m=|M|}^{N-|C|} \tilde{w}_m(s,t)\binom{N}{m}p^m(1-p)^{N-m} \qquad (22)$$

$$K=K_{|M|}+\ldots+K_{N-|C|}$$

denote an estimate of $g(s,t)$ where $\{\tilde{w}_m(s,t);\ m=|M|,\ldots,N-|C|\}$ are computed using algorithm C.2. Then Theorem 2 describes bounds on absolute error.

Theorem 2. For $\tilde{w}_m(s,t)$ and $\tilde{g}_K(s,t)$ algorithm C.2 leads to:

(i) For the finite sequence $\{v_j = \frac{2j-1}{2K_m};\ j=1,\ldots,K_m\}$

$$|\tilde{w}_m(s,t) - w_m(s,t)| \le 2Q_m/K_m \qquad (23a)$$

and

$$|\tilde{g}_K(s,t) - g(s,t)| \le 2\sum_{m=|M|}^{N-|C|} \frac{Q_m}{K_m}\binom{N}{m}p^m(1-p)^{N-m} \qquad (23b)$$

where

$$1 \le Q_m \le \min[R_m(s,t)\ ,\ \binom{N}{m} - R_m(s,t)] \quad m=|M|,\ldots,N-|C|.$$

(ii) There exist infinite sequences $\{v_j\ ;\ j=1,2,\ldots\}$ such that the bounds in (23) hold with $c(\log K_m)/K_m$ replacing $1/K_m$.

The proof of (i) relies on observing that for m operating arcs

$$\tilde{w}_m(s,t) = \frac{1}{K_m}\sum_{j=1}^{K_m} k(s,t|x_j)$$

$$\sum_{\ell\in J_m^*} \frac{1}{K_m}\sum_{j=1}^{K_m} I_{[\frac{\ell-1}{\binom{N}{m}},\ \frac{\ell}{\binom{N}{m}})}(v_j)$$

and on using methods analogous to those employed in the proof of Theorem 1 (i). Part (ii) follows directly.

Choosing a Sample Size

Note that for a given set of sample sizes $\{K_m\}$ the bound in (23b) diminishes to zero as $p \to 0$ and $p \to 1$. However, the choice of $\{K_m\}$ clearly affects this convergence. Observe that

$$Q = \sum_{m=|M|}^{N-|C|} Q_m$$

where Q is defined in (15b). Therefore, if one chooses

$$K_m = K\binom{N}{m}p^m(1-p)^{N-m}/H \qquad m=|M|,\ldots,N-|C|$$

then the error bounds for $|\tilde{g}_K(s,t) - g(s,t)|$ in Theorem 2 and $|\hat{g}_K(s,t) - g(s,t)|$ in Theorem 1 are identical for the finite sequence. For the case of infinite sequences, the dominant term in the bound for $|\tilde{g}_K(s,t) - g(s,t)|$ is identical to the bound for $|\hat{g}_K(s,t) - g(s,t)|$. Moreover, the time complexities for computing $\hat{g}_K(s,t)$ and $\tilde{g}_K(s,t)$ are identical.

An alternative approach for assigning values to K_m makes considerably better use of a priori information, when it is available. Let

λ_m = mean number of steps needed to determine whether or not s and t are connected given that m arcs operate $m=|M|,\ldots,N-|C|$.

Then for a specified number of steps

$$K^* = \sum_{m=|M|}^{N-|C|} K_m \lambda_m , \tag{24}$$

the assignment

$$K_m = K^* \frac{\omega_m}{\sum_{j=|M|}^{N-|C|} \lambda_j \omega_j} \qquad m=|M|,\ldots,N-|C| , \tag{25}$$

where

$$\omega_j = [Q_j \binom{N}{j} p^j (1-p)^{N-j} / \lambda_j]^{1/2} \qquad j = |M|, \ldots, N-|C| \ ,$$

minimizes the bound in (23b). Since $\{Q_m\}$ and $\{\lambda_m\}$ are rarely known, the assignment (25) has limited practical vlaue. Note that van Slyke and Frank chose to omit consideration of the computation time when choosing their assignment of $\{K_m\}$ to minimize variance.

There is an additional feature of stratified sampling that makes algorithm C.2 more appealing than algorithm C.1. Observe that once $\{w_m(s,t) \ ; \ m=|M|,\ldots,N-|C|\}$ is estimated one can use (22) to estimate $g(s,t)$ for as many values of p as desired without the need for additional sampling. By contrast, algorithm C.1 applies for only one value of p at a time. Van Slyke and Frank also mention this advantage for stratified sampling.

3. Unequal Probabilities

This section describes how stratified sampling can be extended to the estimation of $g(s,t)$ with an accelerated convergence rate on the error bound when at least some of the operating probabilities $q_1,\ldots,q_N$ are distinct. Diegert and Diegert (1981) also describe a stratified sampling plan for unequal probabilities, but theirs differs significantly from the plan proposed here both in terms of procedural design and sampling mechanism. In particular, their plan relies on independent replications and random sampling whereas the present proposal uses the specialized finite and infinite sampling sequences, discussed in Section 1, to effect the accelerated convergence rates.

Let

$d_m(s,t)$ = probability that s and t are connected and that exactly m arcs operate.

Observe that

$$d_m(s,t) = \sum_{z \in Z_m} k(s,t|z) \prod_{i \in P} q_i^{z_i}(1-q_i)^{1-z_i} \tag{26}$$

so that

$$g(s,t) = \prod_{i=1}^{N} q_i + \sum_{m=|M|}^{N-1} d_m(s,t) \ . \tag{27}$$

Note the upper limit on summation $N-1$ replaces $N-|C|$ to account for the probabilities $\prod_{i \in P} q_i^{z_i}(1-q_i)^{1-z_i}$ for $N-|C| < \sum_{i=1}^{N} z_i \leq N-1$.

Algorithm C.3 describes a procedure, based on the sampling plan in algorithm C.2, for computing $\bar{d}_m(s,t)$ as an estimate of $d_m(s,t)$. Then our estimate of $g(s,t)$ is

$$\bar{g}_K(s,t) = \prod_{i=1}^{N} q_i + \sum_{m=|M|}^{N-1} \bar{d}_m(s,t) \qquad K=K_{|M|}+\ldots+K_{N-1} \ . \tag{28}$$

Recall that for m operating arcs there are $\binom{N}{m}$ possible combinations of operating arcs and that $J_m=\{1,\ldots,\binom{N}{m}\}$ denotes the set of all indices associated with these combinations. Also, recall that J_m^* denotes the subset of J_m for which s and t are connected when m arcs operate. Now let $z_i(\ell,m)$ denote the status (0 or 1) of arc i on combination ℓ when m arcs operate $i=1,\ldots,N$; $\ell \in J_m$ and $m=|M|,\ldots,N-1$. Then Theorem 3 gives the relevant bounds.

Algorithm C.3

Purpose: To compute $\bar{d}_m(s,t)$ as an estimate of

$d_m(s,t)$ = probability that s and t are connected and exactly m arcs operate.

Given: N, $\{q_i;\ i=1,\ldots,N\}$, m, K_m, s and t. ($|M| \le m \le N-1$).

1. Set parameters: n=N, A=m and S=0.
2. Perform K_m replications:
   a. For $i=1,\ldots,n$ set $x_{ij}=0$.
   b. Select v_j.
   c. Determine the "number" of the combination of A operating arcs and n-A failed arcs:

$$B = \lfloor \binom{n}{A} v_j \rfloor .$$

   d. Determine the A-canonical representation ℓ, $B_\ell,\ldots,B_A$ of B using (13). If B=0, $\ell=A+1$.
   e. If $\ell>1$, then for $v=1,\ldots,\ell-1$; $B_v=v-1$.
   f. Set operating indicators:

      For $v=1,\ldots,A$; set $i=B_v+1$ and $x_{ij}=1$.

   g. Use a labeling algorithm to determine $k(s,t|x_j)$.
   h. Accumulate results:

$$S=S+k(s,t|x_j)\prod_{i=1}^{N} q_i^{x_{ij}}(1-q_i)^{1-x_{ij}} .$$

3. Compute final estimate for K_m replications:

$$\bar{d}_m(s,t) = \binom{N}{m} S/K_m .$$

4. Done.

Theorem 3. Consider $\bar{d}_m(s,t)$, based on algorithm C.3, and $\bar{g}_K(s,t)$ as estimates of $d_m(s,t)$ and $g(s,t)$ respectively. Then

(i) For the finite sequence $\{v_j = \frac{2j-1}{2K_m}\ ;\ j=1,\ldots,K_m\}$

$$|\bar{d}_m(s,t) - d_m(s,t)| \le \frac{2\binom{N}{m}}{K_m} \sum_{\ell\in J_m^*} \prod_{i=1}^{N} q_i^{z_i(\ell,m)} (1-q_i)^{1-z_i(\ell,m)} \tag{29}$$

and

$$|\bar{g}_K(s,t) - g(s,t)| \le 2 \sum_{m=|M|}^{N-1} \frac{\binom{N}{m}}{K_m} \sum_{\ell\in J_m^*} \prod_{i=1}^{N} q_i^{z_i(\ell,m)} (1-q_i)^{1-z_i(\ell,m)} \tag{30}$$

(ii) There exist infinite sequences $\{v_j;\ j=1,2,\ldots\}$ for which (29) and (30) hold with $c(\log K_m)/K_m$ replacing $1/K_m$.

Proof. For (i) note that

$$d_m(s,t) = \sum_{\ell\in J_m^*} \prod_{i=1}^{N} q_i^{z_i(\ell,m)} (1-q_i)^{1-z_i(\ell,m)} \tag{31}$$

whereas

$$\begin{aligned} \bar{d}_m(s,t) &= \frac{\binom{N}{m}}{K_m} \sum_{j=1}^{K_m} k(s,t|x_j) \prod_{i=1}^{N} q_i^{x_{ij}} (1-q_i)^{1-x_{ij}} \\ &= \binom{N}{m} \sum_{\ell\in J_m^*} \Big[\frac{1}{K_m} \sum_{j=1}^{K_m} I_{\left[\frac{\ell-1}{\binom{N}{m}},\ \frac{\ell}{\binom{N}{m}}\right)}(v_j)\Big] \prod_{i=1}^{N} q_i^{z_i(\ell,m)} (1-q_i)^{1-z_i(\ell,m)} \end{aligned} \tag{32}$$

Therefore,

$$|\bar{d}_m(s,t) - d_m(s,t)| = \Big|\binom{N}{m} \sum_{\ell\in J_m^*} \prod_{i=1}^{N} q_i^{z_i(\ell,m)} (1-q_i)^{1-z_i(\ell,m)} \times \frac{1}{K_m} \sum_{j=1}^{K_m} \Big[I_{\left[\frac{\ell-1}{\binom{N}{m}},\ \frac{\ell}{\binom{N}{m}}\right)}(v_j) - \frac{1}{\binom{N}{m}}\Big]\Big|$$

$$\leq \frac{2^{\binom{N}{m}}}{K_m} \sum_{\ell \in J_m^*} \prod_{i=1}^{N} q_i^{z_i(\ell,m)} (1-q_i)^{1-z_i(\ell,m)} \quad ,$$

which establishes (29). The result (30) follows by summation for $m=|M|,\ldots,N-1$. Part (ii) follows directly, completing the proof.

Note that, as in Theorem 2 and 3, the bound (30) diminishes to zero as $p^*=\max_i q_i \to 0$ and $p_*=\min_i q_i \to 1$. However, in contrast to the case of equal q_i , algorithm C.3 requires us to specify the unique $q_1,\ldots,q_N$. Therefore, one needs to use algorithm C.3 for each such set of probabilities considered.

There remains the issue of how to choose $K_{|M|},\ldots,K_{N-1}$. This is not a trivial problem, as Diegert and Diegert show. One possibility is to choose

$$K_m = K \frac{\binom{N}{m}(p^*)^m(1-p_*)^{N-m}}{\sum_{j=|M|}^{N-1} \binom{N}{j}(p^*)^j(1-p_*)^{N-j}} \qquad m=|M|,\ldots,N-|C| \quad . \qquad (32)$$

Then as $p^* \to p_*$ the bound on $|\bar{g}_K(s,t) - g(s,t)|$ in Theorem 3 converges to the bound on $|\hat{g}_K(s,t) - g(s,t)|$ in Theorem 1 for the finite sequence. For the relevant infinite sequences the dominant term in the bound on $|\bar{g}_K(s,t) - g(s,t)|$ converges to the bound on $|\hat{g}_K(s,t) - g(s,t)|$ proportional to $(\log K)/K$. Lastly, the time complexity for computing $\bar{g}_K(s,t)$ converges to that for $\hat{g}_K(s,t)$.

While (32) enables one to achieve the desired accelerated convergence as in the case of equal probabilities in Section 2, better assignments than (32) do exist and conceivably can be identified. But this

identification calls for additional research.

4. An Example

Although the deterministic upper bounds and accelerated convergence rates make apparent the superiority of the proposed method over any alternative method based on independent replications that use random sampling, there remain insights to be gained from an illustration of how these new procedures work in practice. Accordingly, this section illustrates the performance of algorithm C.1 using the van der Corput sequence for estimating $g(s,t)$ in the network of Fig. 1 with $s=1$ and $t=20$. Although slower in convergence than the sequence

$\{v_j = \frac{2j-1}{2K} ; j=1,\ldots,K\}$, the van der Corput sequence enables us to proceed sequentially in the process of achieving a specified accuracy.

Insert Fig. 1 about here.

The network in Fig. 1 has $N=30$ arcs and $|V|=20$ nodes. The minimal cardinality among all minimal cutsets is $|C|=3$ and the minimal cardinality among all minimal pathsets is $|M|=5$. Suppose that all arcs have operating probability p and that the objective is to estimate $g(s,t)$ to within $\pm\delta$. Then for the van der Corput sequence

$$\mathrm{pr}(|\hat{g}_K(s,t) - g(s,t)| \leq \delta) = 1 \qquad (33)$$

for all $K \geq K(\delta)$ where

$$K(\delta) = \min[j: j\delta - 2HQ(\tfrac{1}{3}\log_2 j + 1) = 0;\ j=2,3,\ldots] . \qquad (34)$$

Although the certitude of (33) is appealing its value is, at the moment,

academic since Q and hence $K(\delta)$, are unknown. Therefore we use an alternative approach for studying the performance of algorithm C.1.

Consider a set of 100 independent experiments on the ℓth of which $(\ell=1,\ldots,100)$ the van der Corput sequence (with randomly assigned starting value) is used with algorithm C.1 to compute $\hat{g}_1^{(\ell)}(s,t),\ldots,\hat{g}_{K^*}^{(\ell)}(s,t)$ as estimators of $g(s,t)$, where K^* is a large integer. Also, suppose that on experiment ℓ the quantity

$$K^{(\ell)}(\delta) = \min(j:\ |\hat{g}_i^{(\ell)}(s,t) - g(s,t)| \le \delta \quad j \le i \le K^*) \qquad (35)$$

is computed.

Since there is a global deterministic upper bound $K(\delta)$ for all van der Corput sequences, there must be a finite number of replications associated with each experiment that guarantees the accuracy δ with certainty. If K^* is sufficiently large, say, $K^* \ge K(\delta)$, then $K^{(\ell)}=K^{(\ell)}(\delta)$ $\ell=1,\ldots,100$ are the required numbers of replications for these 100 experiments and a study of the properties of $K^{(1)},\ldots,K^{(100)}$ should provide information about the required number of replications to achieve the accuracy δ with an arbitrary van der Corput subsequence and algorithm C.1.

For our experiments we used $p=.95$, leading to $g(1,20) = 1-.295414 \times 10^{-3}$. This value of $g(1,20)$ is actually a long run estimate based on using algorithm C.1 with $K=2^{22}=4194304$ points generated from the van der Corput sequence. Suppose the objective was to estimate $1-g(1,20)$ to within $\pm\delta$ where $\delta=10^{-5}$. A preliminary analysis of the long run had indicated that our value for $1-g(1,20)$ was well within this bound. Moreover, the long run had also provided evidence

that setting $K^*=2^{21}=2097152$ in (35) was reasonable.

The quantities $K^{(1)},\ldots,K^{(100)}$ were computed on 100 independent experiments each using the van der Corput sequence with a randomly assigned starting value. The sample mean number of required observations was 326469.58 with estimated standard error 21289.28 . The smallest $K^{(\ell)}$ was 63343 , the largest $K^{(\ell)}$ was 857595 and the sample coefficient of variation was 65.211. Note that the largest value was well within our specified K^* . Figure 2 shows the empirical distribution function of the data with the deciles marked.

Insert Fig. 2 about here.

As a basis for comparison, suppose we had used algorithm C.1 with pure random sampling to determine $v_1, v_2,\ldots$. Then Chebyschev's inequality gives for J independent replications

$$\text{pr}(|\hat{g}_J - g| \le \delta) \ge 1 - g(1-g)/J\delta^2$$

where $\hat{g}_J=\hat{g}_J(s,t)$ and $g=g(s,t)$. If δ is sufficiently small so that J is sufficiently large, then

$$\text{pr}(|\hat{g}_J - g| \le \delta) \approx 2\Phi(\delta\sqrt{J/g(1-g)}) - 1$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of the standardized normal distribution.

Suppose we specify a confidence level

$$\text{pr}(|\hat{g}_J - g| \le \delta) = 1 - \alpha \qquad 0 < \alpha < 1 .$$

Then for pure random sampling the approximating normal distribution

leads to a minimal required sample size

$$J(\delta,\alpha) \approx [d(\alpha)/\delta]^2 \, g(1-g) \tag{36}$$

where

$$d(\alpha) = \min[d: \phi(d) = 1 - \alpha/2] \, .$$

For a confidence level $1 - \alpha = .99$ one obtains

$$J(.00001,.01) \approx 19506035$$

from (36), making apparent the considerably larger sample size that random sampling requires to achieve the accuracy $\pm 10^{-5}$ 99 out of 100 times on average.

A FORTRAN program based on algorithm C.1 with the van der Corput sequence used an average of 1.391 microseconds per replication. The same program using a sequence of independent uniform deviates required an average of 1.329 microseconds per replication. Therefore, the van der Corput sequence requires $326469.58 \times .001391/60 = 7.57$ minutes to achieve the desired accuracy whereas independent uniform deviates require $19506035 \times .001329/60 = 432.06$ minutes. The advantage of using the van der Corput sequence for estimating $g(s,t)$ for this network is apparent.

5. <u>(s,T) Connectedness</u>

We now turn to the estimation of

$g(s,T)$ = probability that node s is connected to node t for all $t \in T \subseteq E-\{s\}$.

On replication j the conditional probability of (s,T) connectedness is

$$k(s,T|x_j) = \prod_{t \in T} k(s,t|x_j) \, . \tag{37}$$

Then analogous to (5) for the case of $q_1=\ldots=q_N=p$, this suggests the estimator

$$\hat{g}_K(s,T) = \frac{1}{K} \sum_{j=1}^{K} k(s,T|x_j) \ , \tag{38}$$

if one modifies algorithm C.1 as follows:

(i) Replace the input node t by the node set T .

(ii) Replace $|C|$ and $|M|$ everywhere by unity.

(iii) Replace $k(s,t|x_j)$ by $k(s,T|x_j)$ in steps 2h and 2i.

The motivation for modification (ii) is to account for the fact that there may be different sets $C(s,t)$ and $M(s,t)$ for each $t\epsilon T$ and these may not all be known. If they are known one can replace $|C|$ by $\min_{t\epsilon T} |C(s,t)|$ and $|M|$ by $\min_{t\epsilon T} |M(s,t)|$.

Note that the results in Theorem 1 continue to hold. It is also possible to amend algorithms C.2 and C.3 in a similar manner and have the corresponding properties in Theorem 2 and 3 continue to hold.

References

1. Aho, A.V., J.E. Hopcroft and J.D. Ullman (1974). The Design and Analysis of Computer Algorithms, Addison-Wesley.

2. Ball, M.O. and J.S. Provan (1983). "Calculating Bounds on Reachability and Connectedness in Stochastic Networks", Networks 13, 253-278.

3. Bejian, R. and H. Faure (1977). "Théorie des Nombres - Discrépance de la Suite de van der Corput", C.R. Acad. Sc. Paris, t. 285, sèrie A, 313-316.

4. Chung, K.L. (1949). "An Estimate Concerning the Kolmogoroff Limit Theorem", Trans. Amer. Math. Soc., 67, 36-40.

5. Diegert, C. and K.V. Diegert (1981). "How to Cheat a Monte Carlo Estimation of a System's Reliability", presented at the 1981 Department of Energy Statistical Symposium, Brookhaven National Laboratory.

6. Easton, M.C. and C.K. Wong (1980). "Sequential Destruction Method for Monte Carlo Evaluation of System Reliability", IEEE Transactions on Reliability, R-29, 27-32.

7. Fishman, G.S. and L.R. Moore (1981). "Sampling from a Discrete Distribution While Preserving Monotonicity", The American Statistician, to appear.

8. Fishman, G.S. (1983). "An Alternative to the Monte Carlo Estimation of Network Reliability", UNC/ORSA/TR-83-2, Curriculum in Operations Research and Systems Analysis, University of North Carolina at Chapel Hill.

9. Hammersley, J.M. and D.C. Handscomb (1964). Monte Carlo Methods, Methuen.

10. Karp, R. and M.G. Luby (1983). "A New Monte Carlo Method for Estimating the Failure Probability of an n-Component System", Computer Science Division, University of California, Berkeley.

11. Katona, G. (1966). "A Theorem of Finite Sets", Theory of Graphs, Proc. Tihany Colloquium, September 1966, P. Erdos and G. Katona, eds., 187-207.

12. Kruskal, J.G. (1963). "The Number of Simplices in a Complex", Mathematical Optimization Techniques, R. Bellman, ed., University of California, Berkeley, 251-278.

13. Kuipers, L. and H. Niederreiter (1974). Uniform Distribution of Sequences, John Wiley and Sons.

14. Kumamoto, H., K. Tanaka and K. Inoue (1977). "Efficient Evaluation of System Reliability by Monte Carlo Method", IEEE Transactions on Reliability, R-29, 27-32.

15. Niederreiter, H. (1978). "Quasi-Monte Carlo Methods and Pseudo-random Numbers", Bull. Amer. Math. Soc., 84, 957-1041.

16. Provan, J.S. and M.O. Ball (1983). "Computing Network Reliability in Time Polynomial in the Number of Cuts", Operations Research, to appear.

17. Rosenthal, A. (1977). "Computing the Reliability of Complex Networks", SIAM J. Appl. Math., 32, 384-393.

18. Schmidt, W.M. (1977). Irregularities of Distribution, Tata Institute of Fundamental Research, Bombay.

19. van der Corput, J.G. (1935). "Verteilungsfunktionen I, II", Prok. Akad. Amsterdam, 38, 813-821, 1058-1066.

20. van Slyke, R. and H. Frank (1972). "Network Reliability Analysis: Part I", Networks, 1, 279-290.

21. Zemel, E. (1982). "Polynomial Algorithms for Estimating Network Reliability", Networks, 12, 439-452.

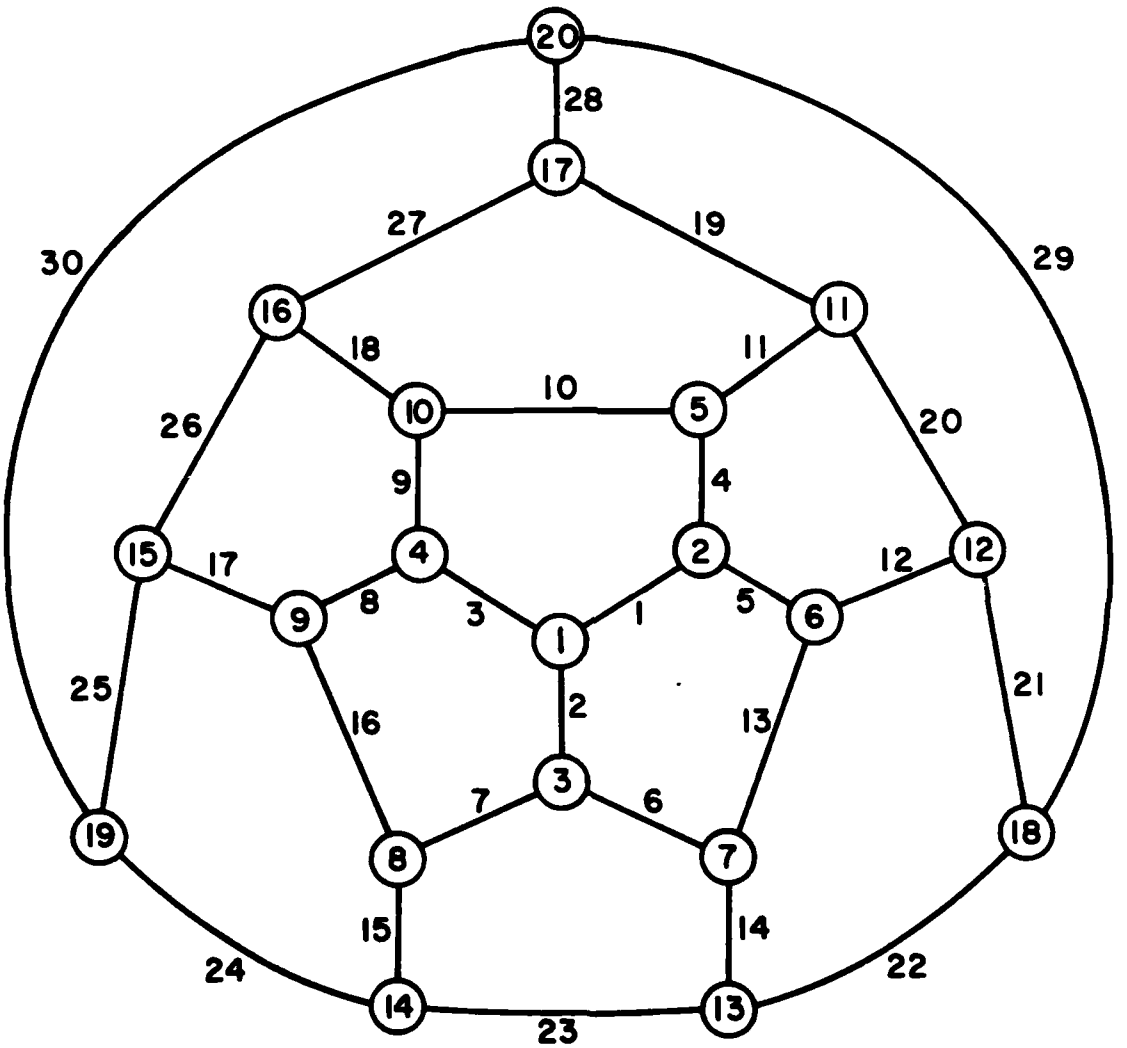


Fig.1 Network

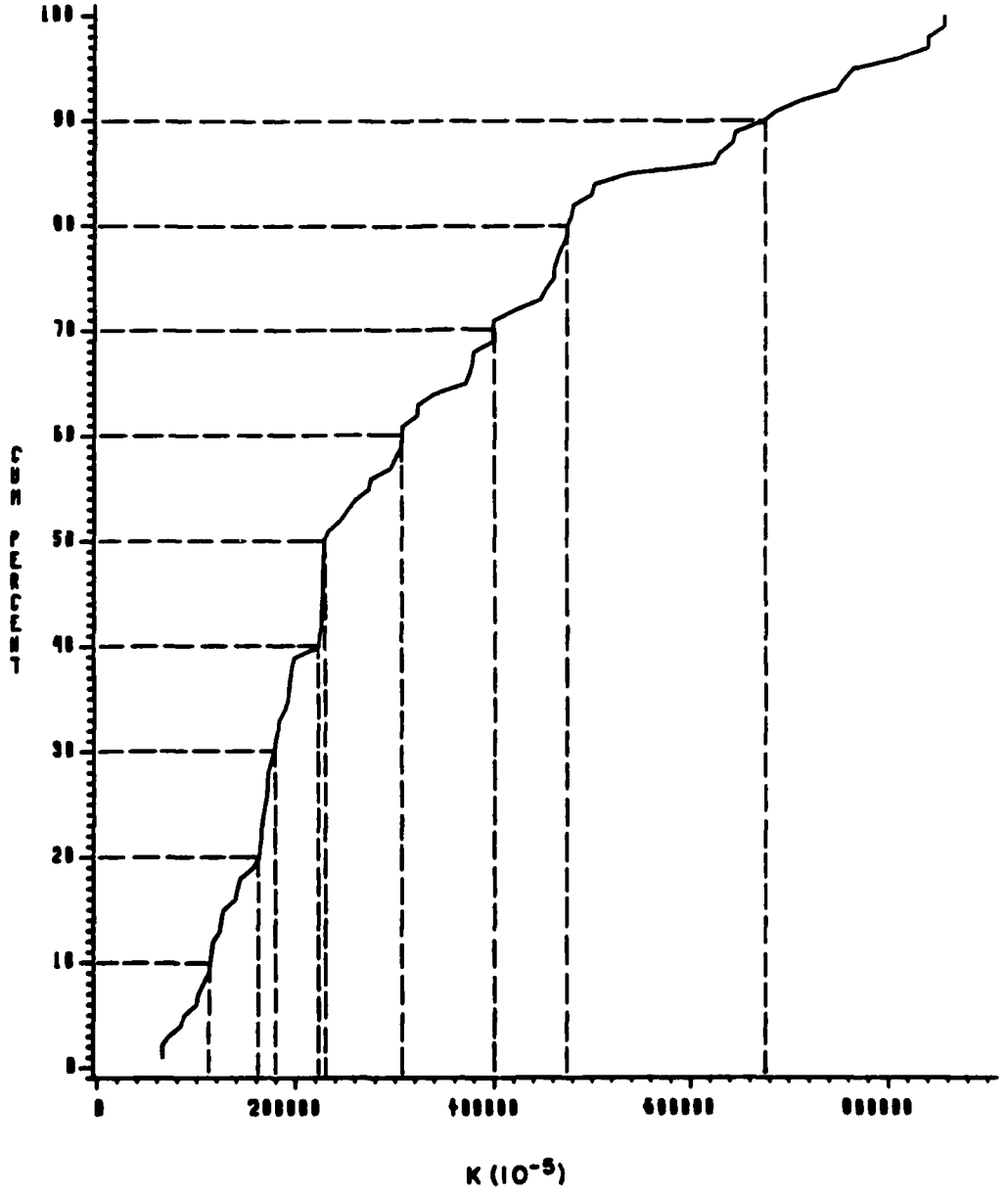


Fig. 2 Empirical Distribution Function for $K(10^{-5})$

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: UNC-ORSA-TR-83/7	2. GOVT ACCESSION NO.: AD-A139841
	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): ESTIMATING NETWORK RELIABILITY WITH ACCELERATED CONVERGENCE RATES ON ERROR BOUNDS	5. TYPE OF REPORT & PERIOD COVERED: TECHNICAL REPORT
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): GEORGE S. FISHMAN	8. CONTRACT OR GRANT NUMBER(s): N00014-76-C-0302
9. PERFORMING ORGANIZATION NAME AND ADDRESS: CURIC. IN OPERATIONS RESEARCH & SYSTEMS ANALYSIS, SMITH BUILDING 128A, UNIV. OF NORTH CAROLINA CHAPEL HILL, NC 27514	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: NAVAL ANALYSIS PROGRAM, OFFICE OF NAVAL RESEARCH, ARLINGTON, VA 22217	12. REPORT DATE: DECEMBER 1983
	13. NUMBER OF PAGES: 34
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report): UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

UNCLASSIFIED

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Let $G=(V,E)$ denote an undirected network with node set V and arc set $E=\{1,\ldots,N\}$. Arcs fail randomly and independently with probability $1-q_i$ for $\forall\ i\in E$. This paper describes a numerical procedure for estimating $g(s,t)$, the probability that nodes s and t are connected for specified $s,t\in E$, with

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE S/N 0102-LF-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

bounds on absolute error proportional to $1/K$ for a specified nonrandom finite sequence and proportional to $\log K/K$ for certain nonrandom infinite sequences, where K is the number of replications. These convergence rates are best possible. Although the infinite sequences have a slower convergence rate than the finite sequence has, they offer the convenience of allowing one to add replications and retain the rate whereas the finite sequence does not. These bounds improve on the convergence rate $O(1/K^{1/2})$ for the standard error in the case of independent Monte Carlo replications based on random sampling. Moreover, they hold with certainty. Algorithms for computing estimates are shown to have an upper bound $O(N)+O(\max(N, V))$ on time complexity per replication as $K \to \infty$.

The paper first describes the estimation of $g(s,t)$ for $q_1=\ldots=q_N=p$ by using a tabled truncated binomial distribution together with the B-canonical representation of a positive integer. It then describes how to incorporate stratified sampling to estimate $g(s,t)$ as a function of p at small marginal increase in time complexity. Next, the paper extends the stratified sampling method to the case of unequal $q_1,\ldots,q_N$.

An example based on a network of 30 arcs illustrates the techniques. Lastly, the paper extends the method to the estimatiof of $g(s,T)$, the probability that s and t are connected for all $t \epsilon T \subseteq E-\{s\}$.